About the Modeling of Radio Source Time Series as Linear Splines

Maria Karbon[1], Robert Heinkelmann[1], Julian Mora-Diaz[1], Minghui Xu[2], Tobias Nilsson[1], Harald Schuh[1]

Abstract Many of the time series of radio sources observed in geodetic VLBI show variations, caused mainly by changes in source structure. However, until now it has been common practice to consider source positions as invariant, or to exclude known misbehaving sources from the datum conditions. This may lead to a degradation of the estimated parameters, as unmodeled apparent source position variations can propagate to the other parameters through the least squares adjustment. In this paper we will introduce an automated algorithm capable of parameterizing the radio source coordinates as linear splines.

Keywords VLBI, radio sources, linear splines, MARS

1 Introduction

Within geodetic VLBI only one global source position for a source's whole VLBI history is estimated. However, many of the observed sources show systematic linear or non-linear behavior and thus influence the stability of the Celestial Reference Frame (CRF), e.g., [1]. Such instabilities, if left unmodeled, will distort the estimated values of the other parameters [2, 3]. To minimize these effects, only sources considered to be stable are entered into the celestial datum definition, the so called ICRF2 (second realization of the International Celestial Reference Frame: [4]) defining sources. Their selection was based on statistical properties of the time series, the observational history, and their position in the sky to guarantee an optimal geometric distribution. However, not all defining sources remain stable over longer time spans, and statistics can be misleading [5, 6]. Thus, we decided to parameterize the source coordinates with linear splines to allow variations, which additionally will allow us to include the so called special handling sources in the datum, making use of the long observational history of these sources.

2 Data and Data Analysis

For our study we used 4,170 VLBI sessions observed within the years 1980 and 2013. We restricted the sessions to those which have a globally distributed station network that includes more than 10^{15} m^3. For each session we estimated the positions of all sources which had more than three observations; otherwise they were constrained to their a priori positions. In the first step the NNR condition for the celestial datum definition was applied to the ICRF2 defining sources. The parameterization for the other parameters was similar to what is described in [7]. From this we were able to get time series of 38 special handling sources and 3,048 sources which are neither defining nor special handling sources, the so called 'other' group. To get the positions of the 265 defining sources within our solution free of datum constraints, we divided the datum sources into two groups (Figure 1). To estimate the positions of one group, the other was included into the NNR condition, and vice versa.

1. GFZ German Research Centre for Geosciences
2. Shanghai Astronomical Observatory

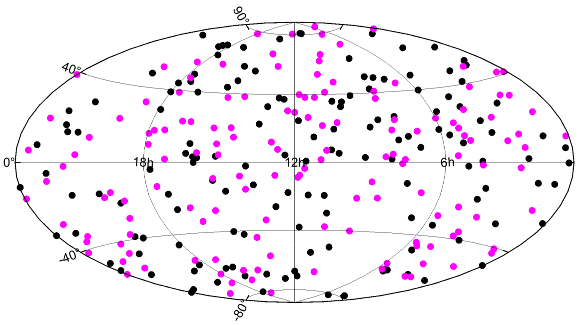

Fig. 1 ICRF2 defining sources, in magenta (light gray) and black, the two groups used for the datum definition to estimate the respective other group.

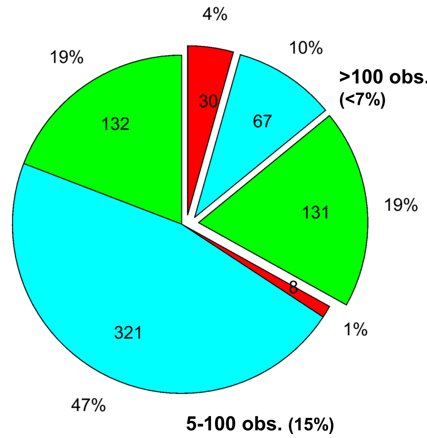

Fig. 2 Number of sources which are observed in 5—100 sessions and more than 100, color coded regarding their class; green (intermediate shading): defining, red (darkest shading): special handling, cyan (lightest shading): other.

3 Time Series of the Sources

80% of the sources contained in our data set are observed in fewer than five sessions, with 2,660 sources falling in the ‘other’ group but also two defining sources being that poorly observed. Figure 2 shows a pie chart of the relative and absolute numbers of sources observed in more than five but less than 100 sessions (15% of the total number of sources), and sources observed in more than 100 sessions (less than 7% out of the total). As one can see, the biggest group (47%) is formed again by sources from the ‘other’ category observed in five to 100 sessions. However, the 131 defining sources which are well-observed (in more than 100 sessions), only amount to 19%, and almost exactly the same number is observed in fewer than 100 sessions. Further, most of the special handling sources are very well-observed, but due to the fact that source position variations are not modeled, they cannot be introduced as datum sources.

Within ICRF2 the assessment of the stability of the source positions was done through statistics, e.g. weighted root mean square, standard deviation, and χ^2 per degree of freedom, while other authors, e.g. [5], use the Allan deviation [8]. That is problematic as these measures depend on the size of the sample, and further, especially in the case of the Allan variance, on the sampling length. The number of sessions per source, or the observation period of a source in years (criteria used for ICRF2), alone do not guarantee that either necessity is fulfilled.

To illustrate that issue, we plotted in Figure 3 three sources which are well-observed; in red (left plots) are $d\alpha$ and $d\delta$ for the special handling source 4C39.25 observed in 3,098 sessions, in green (middle plots) are plots for the defining source 0420-014 observed in 1,032 sessions, and in cyan (right plots) are plots of 0602+673 of the ‘other’ group observed in 759 sessions. In black the semi-annual mean values are plotted. The special handling source is very well-sampled within the whole time span, although more sparsely within the first and last years. Nevertheless, the evolution of the source position variation is well resolved. The defining source for which we have position estimates for the same time span, however, shows huge gaps in the time series. For such a time series the mean values might still give reasonable results as the sample size is big enough. Still, the investigation of spectral characteristics using the whole time span is destined to fail due to the large gaps. This source further shows from the mid ’80s until the mid ’90s clear systematics in the $d\alpha$ coordinate, comparable in magnitude and time span to the special handling source 4C39.25 — a characteristic we do not want for defining sources. The ‘other’ source 0602+673 on the right, however, shows a very stable behavior for both components from 2002 until 2008; then systematics in $d\delta$ emerge.

Considering all of this, we suggest to abandon the classification of sources into special handling and defining sources. Not only is the determination of the stability of the source position highly dependent on the properties of the time series and the statistics applied, but also a source which shows no systematics during one time period can exhibit variations at other

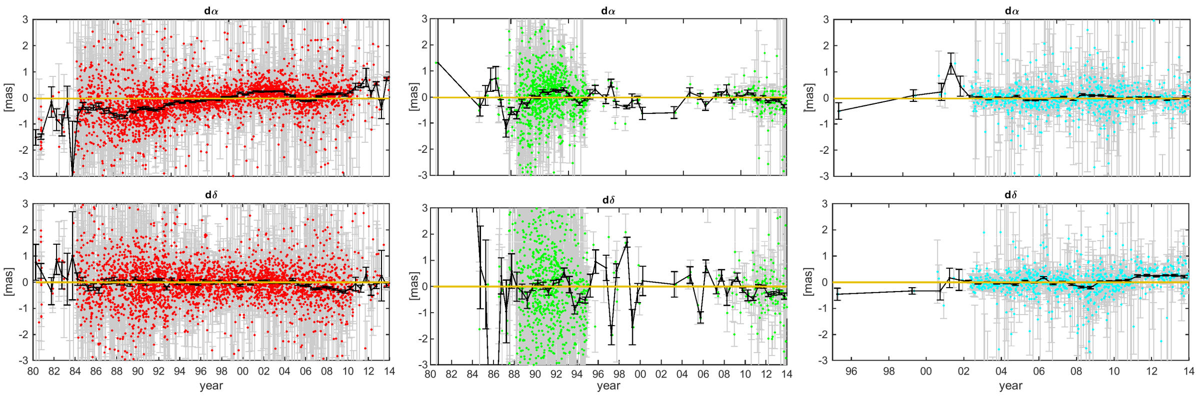

Fig. 3 In red (left plots) are the coordinates of special handling source 4C39.25, in green (middle) are defining source 0420-014's plots, and in cyan (right) are 'other' source 0602+673's plots. All have their error bars in gray. The semi-annual mean values are in black, with corresponding error bars.

times, and vice versa. Also, with increasing precision of the VLBI technique, the demands on the accuracy of the sources are increasing. Thus, holding onto such stability criteria will inevitability lead to a shrinking number of sources suitable for the datum definition. Instead we propose to parameterize the source position variations with linear splines, allowing them the freedom to change — an approach similar to what has been successfully used for years for the station positions.

4 Parameterization of Source Positions with Linear Splines

When parameterizing the source positions with linear splines, one faces the practical problem of the large number of sources. As any manual approach is out of the question when confronted with more than 6,600 time series, we looked for an adequate automated tool. We found what we needed in the MARS algorithm [9], which is a method for flexible regression modeling delivering continuous linear splines. The model consists of a weighted sum of basis functions $B_i(x)$, where each c_i is a constant coefficient.

$$\hat{f}(x) = \sum_{i=1}^{n} c_i B_i(x) \;, \tag{1}$$

Each basis function $B_i(x)$ is a constant, a hinge function, or a product of two or more hinge functions. Hinge functions take the form of

$$max(0, x - const) \quad \text{or} \quad \max(0, \mathrm{const} - \mathrm{x}) \;. \tag{2}$$

To find the ideal set of basis functions, it relies on a fast least squares update technique, trying to minimize the sum-of-squares residual error. The only input data the algorithm needs is the time series and its error information, in our case the single estimates for each source coordinate for each session with its standard deviation.

5 The Linear Splines Determined by MARS

Figure 4 shows the estimates of the source positions of the special handling source 4C39.25 in red (left plots) and estimates of the 'other' source 0602+673 in cyan (right), overlaid with the spline determined by MARS in blue. Both are determined by entering all ICRF2 defining sources into the NNR condition, which is the most robust procedure. As one can see, the MARS spline follows to great extent the semi-annual mean values (plotted in gray). Only where the estimates show larger uncertainties does the algorithm downweight the positions considerably; thus the segmentation of the spline remains unaffected. This can be seen at the beginning of the time series of 4C39.25, where the data is not only more sparse, but also less reliable.

In magenta (lightest thick horizontal line) and black (darkest line) the splines determined through the time series produced by the two different groups of defining sources are plotted. The overall agreement for both sources is good, as expected, although in the first years of 4C39.25 larger differences appear. This is linked to

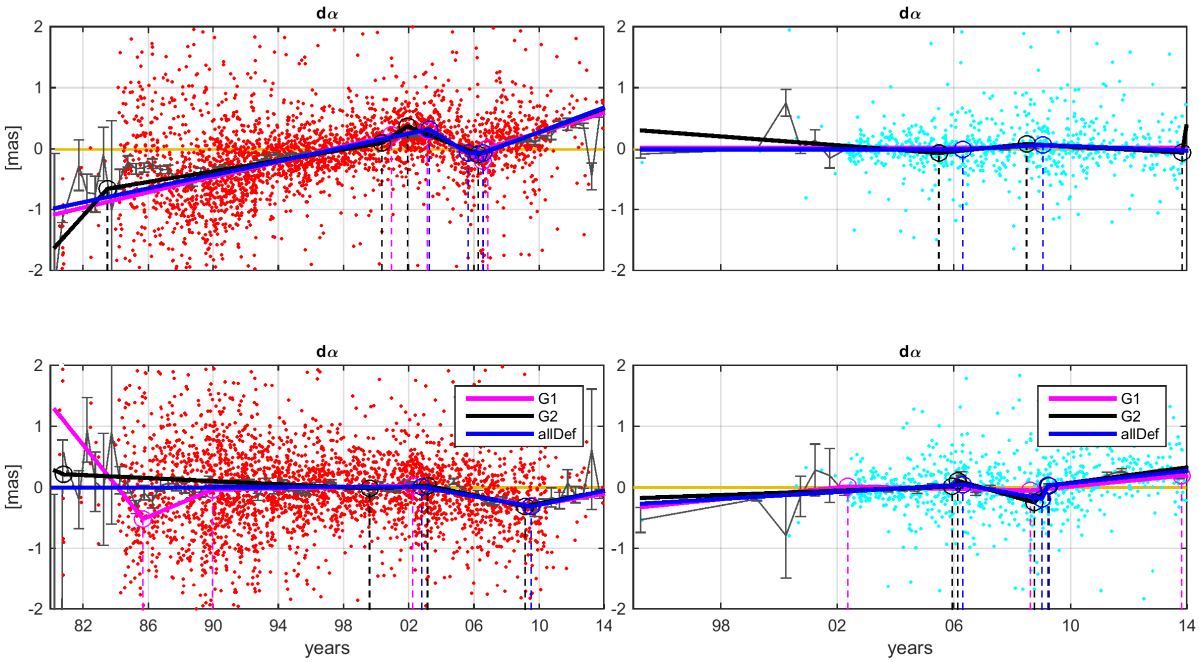

Fig. 4 In red (left) are the coordinates of special handling source 4C39.25, in cyan (right) are the 'other' source 0602+673's coordinates, and in gray are the semi-annual mean values. The time series is determined with all ICRF2 defining sources in the datum, overlaid with the corresponding MARS spline in blue (horizontal line with intermediate shading). The MARS spline when using Group 1 as datum is in magenta (lightest horizontal line) and when using Group 2 is in black (darkest horizontal line).

the generally worse geometry due to the little number of sources observed in the early years of VLBI. Hence, the geometry suffers when halving the datum sources, as it is done within the grouping. A similar effect cannot be seen in the splines determined for source 0602+673, where the splines agree also well when the data is sparse. Here, we look at a source which was introduced to the VLBI observing schedule much later, where the accuracy of the VLBI system was much higher and the geometry, both in the celestial and terrestrial networks, more stable. Anyway, for sources which are not defining, such problems do not appear, as the best datum definition, i.e., all ICRF2 defining sources, is chosen. However, it comes into play when determining the splines for said defining sources.

To assess this effect, we calculated for the defining source 0420-014 an additional solution, where all ICRF2 defining sources are in the NNR condition, except this one. This additional spline is plotted in Figure 5 in cyan (dashed line). The spline which shows the least movement is the one in magenta (lightest horizontal line). This one is produced with the time series determined when Group 1 defines the datum; source 0420-014 is part of this group. This seems to be the network configuration which allows this source the least movement. However, Group 2 (black) and the solution where only this source was excluded from the ICRF2 defining sources (cyan) deliver the exact same spline (alternating, dashed line). This proves that our grouping of the defining sources does not deteriorate the spline determination. The solution where all ICRF2 defining sources are entered into the NNR condition (blue, darker solid line), including 0420-014, meanders between the others, but shows similar features as the black and cyan spline.

6 Conclusions

VLBI analysis can acknowledge the fact that sources have structure that changes over time. Instead of excluding such sources from the datum, which would inevitably lead to a declining number of defining sources as the capabilities of the VLBI system increase, we could model these variations. In this work we present one possible approach to how to parameterize source positions with linear splines. In conclusion, it can be

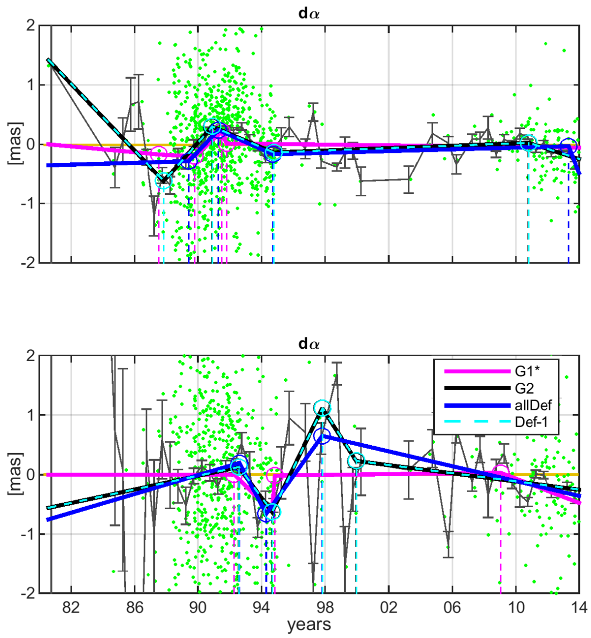

Fig. 5 In green the coordinates of the defining source 0420-014, in gray the semi-annual mean values. Time series determined with Group 2 in the datum (0420-014 is in Group 1); overlaid with the corresponding MARS spline in black (dark part of dashed line). The MARS spline when using Group 1 as datum is in magenta (lighter solid line), when using ICRF2 defining is in blue (darker solid line), and when only excluding this source from the ICRF2 defining is in cyan (light part of dashed line).

said that the crucial aspect of the definition of such splines is the sampling of the data. Whenever the time resolution is good, all the splines, independent of the datum definition and the type of source, agree at a reasonable level. However, whenever the observations get sparse and the reliability of the estimated positions low, the spline depends more on the definition of the datum. Thus, if we want to model the time evolution of the sources the continuous monitoring of VLBI radio sources is indispensable. Results of the application of such splines in the VLBI data analysis are planned to be presented in a paper in the Special Issue of the Journal of Geodesy: VLBI contribution to reference frames and Earth's rotation studies.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft, DFG. Project number: HE 5937/2-1

References

[1] P. Charlot, D. A. Boboltz, A. L. Fey, E. B. Fomalont, B. J. Geldzahler, D. Gordon, C. S. Jacobs, G. E. Lanyi, C. Ma, C. J. Naudet, J. D. Romney, O. J. Sovers, and L. D. Zhang. THE CELESTIAL REFERENCE FRAME AT 24 AND 43 GHz. *Astronomical Journal*, 139(5), 2010.

[2] M. Feissel-Vernier, C. Ma, A. M. Gontier, and C. Barache. Sidereal orientation of the Earth and stability of the VLBI celestial reference frame. *Astron Astrophys*, 438:1141–1148, 2005.

[3] D. S. MacMillan and C. Ma. Radio source instability in VLBI analysis. *J Geod*, 81:443–453, 2007.

[4] A. L. Fey, D. Gordon, C. S. Jacobs, C. Ma, R. A. Gaume, E. F. Arias, G. Bianco, D. A. Boboltz, S. Boeckmann, S. Bolotin, P. Charlot, A. Collioud, G. Engelhardt, J. Gipson, A. M. Gontier, R. Heinkelmann, S. Kurdubov, S. Lambert, S. Lytvyn, D. S. MacMillan, Z. Malkin, A. Nothnagel, R. Ojha, E. Skurikhina, J. Sokolova, J. Souchay, O. J. Sovers, V. Tesmer, O. Titov, G. Wang, and V. Zharov. The Second Realization of the International Celestial Reference Frame by Very Long Baseline Interferometry. *The Astronomical Journal*, 150(2), 2015.

[5] A. M. Gontier, K. Le Bail, M. Feissel, and T. M. Eubanks. Stability of the extragalactic VLBI reference frame. *A&A*, 375:661–669, 2001.

[6] M. Karbon, R. Heinkelmann, Julian Mora-Diaz, Minghui Xu, Tobias Nilsson, and Harald Schuh. About the extension of the parametrization of the radio source coordinates in geodetic VLBI and its impact on the time series analysis. *J Geod*, 2016. submitted.

[7] R. Heinkelmann, T. Nilsson, M. Karbon, L. Liu, C. Lu, J. A. Mora-Diaz, E. Parselia, V. Raposo-Pulido, B. Soja, M. Xu, and H. Schuh. *IVS 2014 General Meeting Proceedings* chapter The GFZ VLBI solution—characteristics and first results. Science Press, Shanghai, China, ISBN 978-7-03-042974-2, 2014.

[8] D. W. Allan. Statistics of atomic frequency standards. *Proc IEEE*, 54, 1966.

[9] J. H. Friedman. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1):1–141, 1991.